DOCUMENT RESUME

ED 138 617 TM 006 238

AUTHOR Brown, Ric; Carbonari, Joseph P.
TITLE Multiple Regression and Canonical Variate Analysis Using Various Modes of Statistical Control.
PUB DATE [Apr 77]
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Correlation; Factor Analysis; Mathematical Models; *Multiple Regression Analysis; Nursing; *Predictive Validity; *Predictor Variables; *Statistical Analysis; Verbal Ability
IDENTIFIERS Canonical Analysis; National League for Nursing Test Battery; Principal Components Analysis

ABSTRACT

Identification and explication of construct relationships, under conditions of extraneous variable control in multiple regression and its multivariate analog, canonical analysis, were studied. Several data models were generated as a function of the interaction of partial correlation and orthogonal linear transformations on nursing examination scores. Multiple regression and canonical analysis procedures were then applied seeking maximum clarity of the underlying relationships. Discussion was concerned with the combinations of analytic techniques that might enhance a researcher's ability to identify and explain salient relationships. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138617

Multiple Regression and Canonical Variate Analysis Using Various Modes of Statistical Control (VAR)

AERA Session 27.14

RIC BROWN
California State University, Fresno

JOSEPH P. CARBONARI
University of Houston

ABSTRACT

Identification and explication of construct relationships, under conditions of extraneous variable control in multiple regression and its multivariate analog, canonical analysis, were the foci of this study. Several data models were generated as a function of the interaction of partial correlation and orthogonal linear transformations on nursing examination scores. Multiple regression and canonical analysis procedures were then applied seeking maximum clarity of the underlying relationships. Discussion was concerned with the combinations of analytic techniques that might enhance a researcher's ability to identify and explain salient relationships.

TM006 238

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Introduction

Construct explanation using multiple regression or its multivariate analog canonical correlation has been hindered by such concerns as fluctuating beta weights or canonical variate weights, suppression effects (Cohen and Cohen, 1975), etc. Also, the investigator is often concerned with the spurious relationships which are a function of extraneous variables such as test form, verbal facility, general intelligence, etc. In seeking explication of domains, or relationships among domains, free of such influences, control of extraneous variables is necessary. Kerlinger (1973) suggests partial correlation which attends to a subtly different research question. This inquiry sought to investigate the clarity of potential solutions under different models of extraneous variable control in regression as well as canonical analysis.

The substantive problem concerned understanding the predictability of the State Board Examinations (SBE) in nursing from the National League for Nursing (NLN) test battery. Predictive validity has been established, but the nature of the relationship is not clear. Previous research has shown the NLN tests to be valid predictors of the SBE (Bell, 1975), but in the same study it was suggested that predictability may be a function of verbal ability being measured in all ten tests rather than content congruity. The NLN was written to serve as a predictor of success on the SBE, therefore, the five topics covered on the NLN are parallel in content to the five topics covered on the SBE. It would be useful for a nursing school to be able to pinpoint

certain problem areas on the SBE that might be remediated during the nursing education process.

Data Source and Techniques

Data for this study consists of a verbal and quantitative score from the Undergraduate Record Exam, scores from the five National League for Nursing exams (NLN), and scores from the five State Board Examinations (SBE) collected on 162 graduates from a Houston school of nursing. The NLN tests (Medical Surgical I, Medical Surgical II, Nursing of Children, Obstetrical, Pediatric and Psychiatric) were written to serve a predictive function and to guide remedial processes.

The following models were examined in an effort to uncover relationships between 2 sets of variables free of a possible contaminant:

1) A multiple regression analysis using the five NLN exams to predict each of the five SBE was conducted.

2) Canonical correlations were calculated using the two sets of variables (SBE and NLN).

3) Both the NLN tests and the SBE (after examination of the variance-covariance matrix) were submitted to a Principal Components factoring procedure with Varimax rotation. The verbal score was used as a marker to assess the existence of a verbal factor.

4) Prior to regression and canonical analysis, respectively, the verbal score was partialed from each of the variables.

5) Prior to regression and canonical analysis, respectively, the apparent verbal factor was partialed from each variable.

6) Full Principal Component analysis was applied to each set of variables after residualizing with respect to the verbal score; followed by both regression and canonical analysis.

7) Full Principal Component analysis was applied to each set of variables after residualizing for the verbal factor; followed by both regression and canonical analysis.

Analysis and Results

The first part of the analysis examined the relationships between the variable sets using observed scores. Table 1 presents each measure on the SBE regressed on the five NLN predictor tests. The NLN was written to serve as a predictor for the SBE, therefore, the tests are parallel in form (i.e., the five topics tested in the SBE are covered by the NLN). Each of the prediction equations does present a relatively high degree of variance accounted for, but the parallel topic in the NLN was not clearly the largest contributor in predicting its SBE counterpart. While SBE Medical has NLN Medical Surgical I and II as largest contributor to the equation, the other equations are not as clear. For example, SBE Pediatric has NLN Medical Surgical II and Psychiatric as largest contributors, while SBE Psychiatric has NLN Obstetrical and Medical Surgical I as its largest contributors. The canonical analysis presented in Table 2 offers a great deal of relationship between sets, but little in the way of understanding or clarification of which NLN tests best predict SBE measures.

A principal components analysis using the verbal score as a marker demonstrated the relatedness of the ten nursing exams with verbal ability. As shown in Table 3, the SBE and NLN measures are highly correlated with the single, rotated component. The verbal score being highly correlated with the component offers evidence that the component may be one of verbal ability.

Partialing techniques were then applied to the nursing exams to observe relationships of the variables free of the influence of the suspected intervening variable. First, the observed verbal score was

TABLE 1

Regression Analysis With Observed Scores

Dependent Variable	Betas of Independent Variables						
SBE	NLN Medical Surgical I	Medical Surgical II	Nursing of Children	Obstet-rical	Psychi-atric	R	R^2
Medical	.30	.31	.04	.07	.19	.68	.46
Surgical	.23	.16	.23	.17	.11	.70	.49
Obstetrical	.08	.28	.24	.22	.23	.76	.58
Pediatric	.09	.27	.15	.17	.21	.75	.56
Psychiatric	.28	.08	.06	.35	.25	.76	.58

TABLE 2

Canonical Analysis With Observed Scores

SBE Set	Canonical Variate Weights	NLN Set	Canonical Variate Weights
Medical	.22679	Medical Surgical I	.09750
Surgical	.21940	Medical Surgical II	.36923
Obstetrical	.18427	Nursing of Children	.12324
Pediatric	.13305	Obstetrical	.25393
Psychiatric	.41010	Psychiatric	.32980

Canonical Correlation = .86009

TABLE 3

Factor Pattern Matrix

From Principal Components Analysis With Varimax Rotation

Variables	Component I
SBE	
Medical	.82
Surgical	.79
Obstetrical	.83
Pediatric	.86
Psychiatric	.79
NLN	
Medical Surgical I	.83
Medical Surgical II	.86
Nursing of Children	.78
Obstetrical	.75
Psychiatric	.87
Undergraduate Record Exam	
Verbal	.86
Quantitative	.38
Percent of Total Variance	63.2

partialed from each of the nursing exams (Table 4A). Notice, that while the multiple correlation coefficients are not quite as large as those in Table 1, clarity of specific predictor is more closely obtained. With the exception of SBE Pediatric, NLN tests corresponding in topic are largest contributors to their respective SBE counterpart. The canonical correlation with verbal score partialed (Table 4B) has the NLN and SBE Psychiatric tests largest in each set of weights. With only one canonical variate arising, calculation of structure coefficients was not undertaken. After calculation of a score on the principal component, that score was partialed from each of the ten nursing measures. Partialing scores on the apparent verbal component (Tables 5A and 5B) did not present as clear a picture as partialing of the verbal score.

A principal components analysis on both the NLN and SBE after partialing the verbal score and verbal component, respectively, yielded similar results. In each set after partialing, five components arose, each individual test loading univocally on a component. With the verbal score partialed, the five SBE components were regressed on the five NLN components. Just as in Table 4A, when partialing the verbal score, clarity of the regression equations was obtained (Table 6A), while the canonical anlysis (Table 6B) revealed little in the way of clarity. The regression and canonical analysis on the two component sets after partialing the apparent verbal component (Tables 7A and 7B) presented very little clarity across sets. Notice that in both analyses with the principal components the total variability accounted for in the regression analysis has as a limit the canonical relation between the sets, slight differences due to the canonical correlation not being fully completed (1 canonical set only).

TABLE 4A

Regression Analysis - Verbal Score Partialled

Dependent Variable	Betas of Independent Variables						
SBE	NLN Medical Surgical I	Medical Surgical II	Nursing of Children	Obstet-rical	Psychi-atric	R	R^2
Medical	.28	.16	.08	.16	.14	.46	.22
Surgical	.08	.19	.06	.18	.04	.40	.16
Obstetrical	.07	.16	.23	.25	.20	.55	.31
Pediatric	.07	.12	.09	.20	.18	.47	.22
Psychiatric	.05	.28	.10	.23	.32	.63	.41

TABLE 4B

Canonical Analysis - Verbal Score Partialled

SBE Set	Canonical Variate Weights	NLN Set	Canonical Variate Weights
Medical	.10	Medical Surgical I	.02
Surgical	.20	Medical Surgical II	.38
Obstetrical	.25	Nursing of Children	.20
Pediatric	.08	Obstetrical	.40
Psychiatric	.69	Psychiatric	.46

Canonical Correlation = .68744

TABLE 5A

Regression Analysis - Verbal Component Partialled

Dependent Variable	Betas of Independent Variables						
SBE	NLN Medical Surgical I	Medical Surgical II	Nursing of Children	Obstetrical	Psychiatric	R	R^2
Medical	.02	.01	.32	.17	.13	.42	.18
Surgical	.06	.13	.15	.05	.18	.33	.11
Obstetrical	.39	.25	.07	.01	.03	.47	.22
Pediatric	.26	.19	.13	.07	.12	.43	.19
Psychiatric	.28	.03	.12	.03	.05	.31	.09

TABLE 5B

Canonical Correlation - Verbal Component Partialled

SBE Set	Canonical Variate Weights	NLN Set	Canonical Variate Weights
Medical	.21	Medical Surgical I	.73
Surgical	.05	Medical Surgical II	.42
Obstetrical	.55	Nursing of Children	.27
Pediatric	.61	Obstetrical	.13
Psychiatric	.38	Psychiatric	.23

Canonical Correlation = .65304

TABLE 6A

Regression Analysis With Principal Components - Verbal Score Partialled

Dependent Variable	Betas of Independent Variables						
SBE Components	NLN Components Medical Surgical I	Medical Surgical II	Nursing of Children	Obstetrical	Psychiatric	R	R^2
Medical	.32	.29	.14	.12	.17	.28	.08
Surgical	.22	.29	.16	.09	.11	.33	.11
Obstetrical	.19	.22	.21	.31	.12	.24	.06
Pediatric	.07	.21	.28	.23	.19	.33	.11
Psychiatric	.22	.24	.07	.18	.32	.36	.13

TABLE 6B

Canonical Analysis With Principal Components - Verbal Score Partialled

SBE Set Components	Canonical Variate Weights	NLN Set Components	Canonical Variate Weights
Medical	.34	Medical Surgical I	.21
Surgical	.27	Medical Surgical II	.19
Obstetrical	.18	Nursing of Children	.38
Pediatric	.31	Obstetrical	.31
Psychiatric	.30	Psychiatric	.26

Canonical Correlation = .631

TABLE 7A

Regression Analysis With Principal Components - Verbal Component Partialled

Dependent Variable	Betas of Independent Variables						
SBE Components	NLN Components: Medical Surgical I	Medical Surgical II	Nursing of Children	Obstetrical	Psychiatric	R	R^2
Medical	.24	.03	.01	.29	.08	.30	.09
Surgical	.25	.13	.20	.19	.18	.34	.12
Obstetrical	.08	.06	.10	.24	.09	.22	.05
Pediatric	.16	.09	.21	.09	.22	.33	.11
Psychiatric	.10	.22	.30	.01	.16	.36	.13

TABLE 7B

Canonical Analysis With Principal Components - Verbal Component Partialled

SBE Set Components	Canonical Variate Weights	NLN Set Components	Canonical Variate Weights
Medical	.07	Medical Surgical I	.31
Surgical	.28	Medical Surgical II	.17
Obstetrical	.28	Nursing of Children	.18
Pediatric	.11	Obstetrical	.16
Psychiatric	.31	Psychiatric	.22

Canonical Correlation = .656

Discussion

With the assumption that the subtests of the National League for Nursing (NLN) exams ought to predict best their counterparts in the State Board Examinations (SBE) as a criteria, the attempt was to find the best combination of statistical control and analysis method. It was found that partialing a verbal score, the verbal subtest of the Undergraduate Record Exam (UGRE), from either the subtest scores or the principal components derived from the subtests best deliniated the relationships sought.

The verbal score had two attributes that might have been operating within this paradigm. One is the time at which the UGRE was taken. It was prior to, and apart from the NLN. The other is the nature of what the verbal section of the UGRE measured. It is a measure of a global verbal ability, similar to a general verbal factor. However, the courses taken in nursing school should contain, as a part of the "knowledge" of the course, a vocabulary and a level of verbal facility which is specific to the course and to that portion of the nursing profession. This specificity is analogous to a specific verbal factor. Therefore, it is quite likely that when the verbal score was used in the partial correlation, individual differences on general verbal ability were controlled. leaving differences in specific verbal ability a major contributor to the variance-covariance matrix.

When a verbal factor was used, however, both the variance due to general verbal ability and the variance due to specific verbal ability are partialed from the dependent variable set. What systematic variance is left in the matrix is probably a function of some other element

of the tests, such as some form of higher order reasoning, i.e., problem solving, evaluating or synthesizing.

It would be most probable that the degree or amount of higher order reasoning called for in the subtests of the two exams is not congruent with the content area of the subtests. For example, it may be that the Obstetrical subtest of the SBE and the Medical Surgical I subtest of the NLN may both have items in them which call for such reasoning, while the items of the Obstetrical portion of the NLN does not require this higher level skill. Such a situation would account for the finding, for instance, that when the verbal factor was partialed, the Medical Surgical I subtest of the NLN became the biggest contributor to the prediction of the Obstetrics subtests in the SBE.

The nature of the results of the principal components analysis give credence to this explanation. One test only, loaded on each component regardless of whether the verbal score or the verbal factor was partialed prior to the components analysis. Apparently each subtest within the SBE and each within the NLN had a unique component, beyond a general verbal ability and a specific verbal ability that correlated across exams.

The evidence of this study points to the necessity of identifying, a priori, which source of variance the researcher actually wishes to control and identifying the measure or factor that best represents this source. There is a danger that one may remove a part of the variance that is critical to the content. For example, if specific verbal ability is an integral part of knowing a field such as obstetrics, partialing a general verbal ability (as measured by some general test) is the better

approach. If, on the other hand, the researcher is seeking to isolate a relationship that is more completely free of any influence of verbal ability, partialing a verbal factor is the better strategy.

The multiple regression models gave a clearer picture of the relationships of the elements of one test to the elements of the other regardless of the partialing strategy used. This was not unexpected; the canonical models are better descriptors of the overall relationships between domains and the number of dimensions needed to capture the nature of the relationship. Therefore, if a researcher wishes to predict performance on a series of specific tests, multivariate multiple regression might be the better analytic tool. The researcher should be aware, however, that the sum of the R^2s, when subtests are used as dependent variables, is not a valid index of the total relationship unless the intercorrelations among all the dependent variable subtests are zero. In either case, F tests, using a common alpha level such as 0.05 on each regression equation, would result in a inflated experiment-wise error rate.

If the researcher wishes to explain the variance in a construct in one set by predicting it from subtests in the other set, principal components or some from of factor analysis should be used on the dependent variable set. If only orthogonal rotations are used, the sum of the R^2s over the regression models is a valid index of the total relationship of the factored set to the predictor set. F tests on each regression equation, with alpha distributed over the tests by some method such as a Boniferroni process, would be feasible.

The researcher may desire to relate constructs in one set to

constructs in the other. Principal components or factor analysis after any desired partialing of each set would produce orthogonal factors representing the underlying constructs. (Again assuming any rotations done would be orthogonal.) Regression analysis at this point would clearly isolate the independent contribution of each variable in the predictor set to the overall prediction of any variable in the criterion set. It would be indexed by the r^2 between the criterion factor and the predictor factor, because the overall R^2 is equal to the simple sum of the squared bivariate correlations when the predictor variables are uncorrelated.

If used with caution, combinations of analytical techniques can considerably enhance a researcher's ability to identify the salient features of relationships. This, in turn, should more quickly lead to better theoretical models and more easily understood processes.

REFERENCES

Bell, J. A cross-validation study for predictors of scores on the State Board Examinations. Nursing Research, in press.

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlational Analysis for the Behavioral Sciences. New York: John Wiley and Sons, 1975.

Finn, J. D. A General Model for Multivariate Analysis. New York: Holt, Rinehart and Winston, Inc., 1974.

Kerlinger, F. N., and Pedhazur, E. J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1974.

Rummel, R. J. Applied Factor Analysis. Evanston: Northwestern University Press, 1970.